DOCUMENT RESUME

ED 214 155 CS 006 579

AUTHOR Fuchs, Lynn; And Others
TITLE Reliability and Validity of Curriculum-Based Informal Reading Inventories.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Office of Special Education (ED), Washington, D.C.
REPORT NO IRLD-RR-59
PUB DATE Oct 81
CONTRACT 300-80-0622
NOTE 41p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Education; *Informal Reading Inventories; Reading Comprehension; *Reading Instruction; *Reading Research; Reading Tests; *Testing Problems; *Test Reliability; *Test Validity; Word Recognition

ABSTRACT

A study was conducted to explore the reliability and validity of three prominent procedures used in informal reading inventories (IRIs): (1) choosing a 95% word recognition accuracy standard for determining student instructional level, (2) arbitrarily selecting a passage to represent the difficulty level of a basal reader, and (3) employing one-level floors and ceilings of performance to demarcate levels beyond which behavior is not sampled. Subjects were 91 elementary school students, representing a range of reading abilities. The students completed word recognition and passage comprehension tests and then individually read passages from each of the ten reading levels in the Ginn 720 and the nine levels of the Scott-Foresman Unlimited reading series. Correlational and congruency analyses of the resulting data supported the validity of the 95% word recognition accuracy standard, but raised questions about the reliability and validity of the passage sampling procedures and the use of one-level floors and ceilings of performance. The findings suggest that IRI procedures for selecting passages from basal readers and for sampling students' performance at instructional levels may have a negative effect on educational practice. Sampling over time and test forms is a more valid IRI procedure. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214155

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
X Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

University of Minnesota

Research Report No. 59

RELIABILITY AND VALIDITY OF CURRICULUM-BASED INFORMAL READING INVENTORIES

Lynn Fuchs, Douglas Fuchs, and Stanley Deno

IRLD

Institute for Research on Learning Disabilities

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS006519

IRLD

Director: James E. Ysseldyke

Associate Director: Phyllis K. Mirkin

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.

Research Report No. 59

RELIABILITY AND VALIDITY OF CURRICULUM-BASED INFORMAL READING INVENTORIES

Lynn Fuchs, Douglas Fuchs, and Stanley Deno

Institute for Research on Learning Disabilities

University of Minnesota

October, 1981

Abstract

Informal Reading Inventories (IRIs) are endorsed frequently by textbook authors and teacher trainers. However, the reliability and validity of standard and salient IRI procedures rarely have been investigated. Employing 91 elementary age students, this study examined the technical adequacy of (a) choosing a criterion of 95% accuracy for word recognition to determine an instructional level, (b) selecting arbitrarily a passage to represent the difficulty level of a basal reader, and (c) employing one-level floors and ceilings to demarcate levels beyond which behavior is not sampled. Correlational and congruency analyses supported the external validity of the 95% standard but questioned the reliability and validity of passage sampling procedures and one-level floors and ceilings. Sampling over occasions and test forms is discussed as a more valid IRI procedure.

Reliability and Validity of Curriculum-Based Informal Reading Inventories

Certain norm-referenced tests possess strong technical adequacy. Their reliability, together with their capacity to compare the performance of an individual pupil to the performance of a group of similar students, makes them both well suited as instruments for screening and, in some instances, useful for placing pupils in special programs (Salvia & Ysseldyke, 1981). Most normative measures, however, do not have adequate content validity; standardized test items infrequently reflect the content of curricula employed in classrooms (Armbruster, Stevens, & Rosenshine, 1977; Eaton & Lovitt, 1972; Jenkins & Pany, 1978). Thus, normative tests have limited utility for placing pupils in specific instructional programs.

Many years ago, educators with an interest in reading instruction recognized the disparity between the content of standardized tests and the content of classroom curricula. Awareness of this incongruency fueled efforts, such as those by Wheat in 1923, to construct informal reading devices that would be more sensitive to classroom instruction and thereby would be more accurate in assessing students' strengths and weaknesses and their instructional levels (Beldin, 1970).

Curriculum-based Informal Reading Inventories (IRIs) represent one such alternative to normative tests for assessing students' reading behavior. While the extent to which they are employed by classroom teachers is unclear, they are frequently and strongly endorsed by textbook authors and teacher trainers (e.g., Lowell, 1970). Kelly (1970)

typified many academicians' admiration of IRIs when he wrote: "Reading authorities agree that the informal reading inventory represents one of the most powerful instruments readily available to the classroom teacher for assessing a pupil's instructional reading level" (p. 112).

In spite of, or perhaps because of, this popularity, the soundness of procedures that typically govern the use of curriculum-based IRIs rarely has been investigated. This apparent lack of concern may be handicapping educators' efforts to determine accurately students' instructional levels. Evidence for this is provided in occasional studies that investigated the reliability of IRI procedures.

Procedures for Sampling IRI Passages

One prominent feature of curriculum-based IRIs is the procedure of selecting passages by drawing arbitrarily from texts (Beery, Barrett, & Powell, 1969; Bush & Huebner, 1970; Johnson & Kress, 1969). The adequacy of this sampling procedure rests on the assumption that passages are likely to be representative of the texts from which they were selected.

The correctness of this presumption has been questioned indirectly. Investigations have established that extreme variation exists in the readability of basal readers. Not only is there great divergence among basal readers of equal grade designations from different series (Pikulski, 1974), but also there is dramatic variation in passages within the same text (Bradley & Ames, 1977; Fitzgerald, 1980). Such variation suggests that the practice of representing a book's readability level with arbitrarily drawn samples is inadequate, and that this practice may lead to inappropriate instructional placements.

Ceilings and Floors on Performance

While the foregoing concern questions the precision with which

passages represent the difficulty of basal readers, a second concern deals with the adequacy with which curriculum-based IRI procedures sample students' reading skills.

Typically, the first level at which a student fails to meet a criterion of mastery is designated the pupil's "ceiling," and there is no further assessment of reading behavior at levels of difficulty beyond this point. Similarly, reading behavior is not assessed below the level at which a pupil first reads proficiently. This level is designated the student's "floor." The belief that assessment is unnecessary below the one-level floor and above the one-level ceiling rests upon at least two important assumptions. The first is that the difficulty of a series of basal passages progresses steadily so that levels above a ceiling and below a floor represent, respectively, advanced selections and mastered material. This assumption, as discussed above, appears shaky. Second, given materials that are graduated accurately in difficulty, it is assumed that a consistent, inverse relationship exists between the quality of reading behavior and passage difficulty, so that as the difficulty levels of successive passages increase, the reading performance of a student necessarily worsens. Despite the importance of this second assumption to the use of ceilings and floors within IRIs, no pertinent empirical investigations have been identified.

Criteria for Instructional Levels of Performance

In addition to the questionable or unknown reliability of practices that direct the sampling of reading materials and the sampling of reading behaviors, a third prominent feature of IRIs further obscures

the usefulness of the informal reading assessment strategy. This third component is the criterion chosen to determine pupils' levels of reading instruction.

There is no widespread consensus on standards to use for the identification of a pupil's instructional level (Kender, 1969). Traditional criteria in evaluating word accuracy and comprehension are 95% and 75%, respectively. The popularity of this convention, attributed to Betts (Beldin, 1970), is suggested by its use in inventories developed by Harris, Botel, Kress, and Johnson, and Austin and Huebner (Powell, 1971). However, departures from Betts' standards have been numerous and, in some cases, dramatic. Smith (1959), for example, employed a criterion of 80% for word accuracy and 70% for comprehension. Cooper (1952) suggested 96% and 60% as criteria for word accuracy and comprehension, respectively, in the primary grades, and 98% for word accuracy and 70% for comprehension in the intermediate grades. Spache (cited in Lowell, 1970) employed 60% and 75% as satisfactory lower limits of performance for word accuracy and comprehension, respectively.

More important than the lack of agreement on the usefulness of Betts' standards is the indication that the 95% word recognition criterion may have weak internal validity. According to Powell (1971), its possible incorrectness is indicated in two ways. First Killgallon's data, on which the Betts convention is based, appear insufficient in that (a) they represent the performance of only 41 fourth grade students, and (b) the interpretation of subjects' scores was gratuitous.

Second, Powell demonstrated that first and second graders could tolerate an average word recognition score of only 85% and still

maintain 70% comprehension. Pupils in grades 3 through 6 could achieve 70% comprehension with an average word accuracy performance of 91% to 94%. Thus, regardless of grade level, the 95% word recognition criterion was not supported. This finding has received corroboration from Pikulski (1974).

In addition to the questionable internal validity of Betts' standards, persuasive evidence of their external validity is lacking (Kender, 1969). Few studies have attempted to validate the traditional criteria for word accuracy and comprehension against external standards, and available investigations disagree in their findings.

Three studies exemplify this last point. Oliver and Arnold (1978) found that the Iowa Test of Basic Skills (ITSB) correlated more strongly than did the Goudy IRI with teacher judgments concerning the instructional placements of students. Arnold and Arnold (1966) obtained similar results using a curriculum-based IRI, the Gates-MacGinitie Reading Tests, and the Wide Range Achievement Test. However, Botel (1968) found that the Botel Reading Inventory had higher correlations with pupils' actual instructional levels than did the California Reading Test, ITBS, and STEP.

Any conclusions that may be drawn from these conflicting findings become even more tentative in light of several methodological problems in the studies. All of the studies used achievement tests of questionable psychometric adequacy (cf. Ysseldyke, 1979). Also, the studies of Arnold and Arnold (1966) and Oliver and Arnold (1978) used (a) teacher judgments about the placement of pupils for instruction rather than the teachers' actual placements of students, and (b) small samples that pre-

cluded reliable correlations (Nunnally, 1959). Therefore, the instructional performance standard traditionally employed in IRIs lacks both external and internal validity.

In summary, with their high content validity, many curriculum-based IRIs are strong precisely in a way in which most norm-referenced tests are weak. Alternately, however, salient IRI procedures have yet to demonstrate the high degree of reliability that characterizes some standardized instruments. This remains so despite the frequency with which IRIs have been advocated by textbook authors and teacher trainers. The purpose of the present study was to explore the reliability and validity of the three prominent IRI procedures discussed above. This exploration was undertaken not to contribute to the elimination of IRIs but rather to clarify the legitimacy of their use or to strengthen the manner in which they are employed. Specifically, the study (a) explored how many sample passages from basal textbooks were required before the readability levels of the passages represented the readability levels of the textbooks, (b) investigated the consistency of the relationship between pupils' reading performance and passage level difficulty to ascertain the adequacy of current practices that establish floors and ceilings of performance, and (c) examined an array of word recognition criteria to determine which standards, if any, demonstrated acceptable external validity with respect to achievement tests and teacher placements for instruction.

Method

Subjects

Subjects were 91 students (51 boys and 40 girls) randomly selected

from one public elementary school in a metropolitan school district in the Midwest. The numbers of subjects in grades 1-6, respectively, were 14, 17, 15, 18, 16, and 11. Fifteen subjects (16%) participated in a special education resource program, and another 23 subjects (25%) were enrolled in a Title I program for students who had been designated by their teachers as seriously behind in reading.

Measures

Achievement tests. Two tests were selected from the Woodcock Reading Mastery Tests (WRMT)--Word Identification (WI) and Passage Comprehension (PC). The WI test requires that students read aloud isolated words. There are 150 words ranging in difficulty from preprimer to beyond 12th grade level (Woodcock, 1973). The PC test contains 85 items that employ a modified cloze procedure (Bormuth, 1969). Pupils are asked to read silently a passage from which a word has been deleted and to produce verbally an appropriate missing word. The passages range in difficulty from first grade to college level (Woodcock, 1973).

Teacher placements. The classroom teacher of each student reported the book level in the Ginn 720 reading series from which the pupil read for instructional purposes.

Basal readers. Two basal reading series were employed, Ginn 720 (1976) and Scott-Foresman Unlimited (1976). They were chosen as exemplars of popular and contrasting approaches to reading instruction. Ginn 720 emphasizes a combination of phonetic, linguistic, and structural skills, whereas Scott-Foresman Unlimited places primary emphasis on comprehension and study skills.

Procedure

Before testing. Two 100-word passages were selected as measures from each of 10 reading levels in Ginn 720 and 9 reading levels in Scott-Foresman Unlimited. To ensure that these passages were representative of the reading difficulty of the levels from which they were chosen, the following procedure, adapted from Fuchs and Balow (1974), was employed. First, five pages were chosen at random from (a) the last 25% of the pages constituting each reading level, and (b) pages that were not dominated by phonics exercises, dialogue, indentations, and proper nouns. Second, on each of these five pages a 100 word passage was identified. Next, for each passage a readability score was calculated. The Spache Readability Formula (Spache, 1953) was applied to passages in books from preprimer through third grade and the Dale-Chall Formula for Predicting Readability (Dale & Chall, 1948) was used for passages in books from fourth grade through sixth grade. Fourth, the average readability of the five passages at each reading level was determined. Last, if the readability scores of two passages were within one month of the mean readability score of the five passages, then these two passages were selected as representative of that level. However, if two passages could not be identified, then a sixth passage was randomly chosen and steps two through five were repeated. This procedure was repeated until two appropriate passages were found.

Also preceding assessment, classroom teachers indicated the reading level to which each subject was assigned for classroom instruction.

During testing. Subjects individually were administered the

WI and PC tests and were asked to read passages from each of the 10 reading levels in the Ginn 720 and the 9 levels in the Scott-Foresman Unlimited series. This was accomplished in one 45 to 60 minute session in the subject's home school. Testing was conducted by trained research and psychometric assistants.

The reading passages from the basal readers were administered in random order. Preceding the presentation of the first passage, the examiner said, "I want you to read aloud to me as quickly as you can. If you don't know a word, skip it. Try your hardest and remember to read quickly. I'll tell you when to stop." The examiner then presented a copy of the passage, directed the subject to begin, and activated a stopwatch. Subjects were permitted 60 seconds in which to read each passage. The examiner scored each subject's performance by crossing out insertions, substitutions, mispronunciations, and omissions. For each passage, three scores were generated for the subject: the number and percentage of words read correctly and the number of words read incorrectly. For subjects who completed reading a passage in less than the allotted time, the time (in seconds) required by the subject was indicated.

Following testing. Seven criteria were used for judging instructional levels in each of the two reading series. The criteria are defined below. For each criterion, an instructional level was assigned to each subject by identifying the highest reading level at which the subject met the standards before unsatisfactory performance was demonstrated at two consecutive levels.

Criterion 1: for Pre-Primer (PP) through grade 3 books, 30-49 words per minute (wpm) with seven or fewer errors per minute (epm); for grade 4 through grade 6 books, 50 or more (+) wpm with seven or fewer epm.

Criterion 2: 70 + wpm with 10 or fewer epm.

Criterion 3: 100 + wpm with 0-2 epm.

Criterion 4: 95% accuracy.

Criterion 5: 70 wpm with 95% accuracy.

Criterion 6: for PP through grade 2 books, 50 + wpm with 95% accuracy; for grade 3 through grade 6, 70 + wpm with 95% accuracy.

Criterion 7: for PP through grade 2 books, 50 + wpm with 85% accuracy; for grade 3 through grade 6 books, 70 + wpm with 95% accuracy.

Criteria 1-3 were selected because they are employed frequently by precision teachers (Alper, Nowlin, Lemoine, Perine, & Bettencount, 1973; Haughton, 1973; Starlin, 1979; Starlin & Starlin, 1974). Criterion 4 was chosen because it is the traditional standard among users and advocates of IRIs for identifying pupils' instructional levels (Beldin, 1970). Criteria 5 and 6 were devised for this study, and represent combinations of the rate and percentage-accuracy criteria found in the first three criteria. In Criterion 7, an 85% accuracy standard for students in books PP-2 was introduced. Its selection was based on Powell's (1971) demonstration that PP through grade 2 readers maintained 70% comprehension while their word recognition accuracy was at 85% or better.

Results

Representativeness of Sample Passages

Table 1 displays the reading levels from the Ginn 720 and Scott-Foresman Unlimited series and corresponding readability scores both as

reported by publishers and as derived from readability formulae. As shown in Table 1, means of the scores produced by readability formulae were calculated (a) across the total number of passages sampled at each reading level, and (b) on the two 100 word passages at each reading level that were used as measures in the study. Additionally, Table 1 displays the number of passages sampled at each reading level before the readability scores of two passages coincided with the mean readability scores for the reading levels. The number of passages necessary to achieve adequate representation ranged from 5 to 14. Of 19 textbooks in both reading series, 10 (53.00%) required the selection of 10 or more passages before two representative passages could be identified.

Insert Table 1 about here

Difficulty of Passages and Variability of Performance Across Reading Levels

Increasing passage difficulty. Within the two basal series, the mean readability scores of adjacent levels were compared. Differences between pairs of scores, as well as the values of the *t* tests, are presented in Table 2. These contrasts indicate that, for both basal series, the readability scores of the passages increased steadily at successively higher book levels. In Ginn 720, readability scores increased an average .44 grades; in Scott-Foresman Unlimited, scores increased an average .43 grades. Seven of the nine contrasts for Ginn 720 were statistically significant. In Scott-Foresman Unlimited, only three of the eight comparisons were significant. This suggests greater reliability for the differences between adjacent levels in the Ginn 720

series than in the Scott-Foresman Unlimited series. However, given nearly identical increases in readability scores in the two basal series ($\bar{X}$=.44 grades for Ginn 720; $\bar{X}$=.43 grades for Scott-Foresman Unlimited), this greater reliability seems to be due to reduced variability in the readability of passages in Ginn 720 rather than to larger differences in the readability scores between selected passages.

Insert Table 2 about here

Variability of student performance. Two analyses were employed to determine whether performance decreased as the difficulty of sample passages increased. The first analysis examined the group's mean performance on increasingly more difficult passages.

Figure 1 displays mean words correct per minute (wpm), mean errors per minute (epm), and mean percentage correct (pc) scores in both basal series. Trend lines (White, 1971) were computed on and drawn through the data in Figure 1. The trend lines revealed a negative slope for mean wpm scores (-5.33 in Ginn 720 and -2.56 in Scott-Foresman Unlimited) and for mean pc scores (-3.50 in Ginn 720 and -.88 in Scott-Foresman Unlimited). As expected, the mean performance scores generally decreased as passage difficulty increased. However, this was not a consistent performance pattern. Of 17 pairs of adjacent passages that increased in difficulty, 13 pairs (76.00%) of mean wpm scores and only 11 pairs (65.00%) of mean pc scores decreased. This inconsistency in performance is more obvious with respect to the mean epm scores. While the trend line for Ginn 720, as anticipated, was positively sloped (+.89), the

trend line for Scott-Foresman Unlimited was flat. Moreover, among the 17 pairs of sample passages that increased in difficulty, only 9 pairs (53.00%) of mean epm scores increased.

Insert Figure 1 about here

Standard deviations of the mean scores plotted in Figure 1 ranged from 47.8 to 37.5 for wpm scores, 31.6 to 39.0 for pc scores, and 9.9 to 20.7 for epm scores. Given this variability, a congruency analysis was undertaken to explore the regularity with which each subject's performance reflected sample passages' increasing difficulty. An index of the degree of variability of subjects' performance, calculated for each instructional criterion and for both series, was defined as the percentage of subjects (a) failing to meet the instructional criterion at a level lower than the one where that criterion had been met successfully, and/or (b) meeting the instructional criterion at a level higher than one at which the criterion already had been failed. Averaged across the seven instructional criteria and the two basal series, 55.00% of the subjects showed this inconsistency in performance. For the traditional IRI standard, 95% accuracy of word recognition, 56.00% of the subjects demonstrated this inconsistency.

Validity of Alternative Instructional Criteria

Correlational and congruency analyses were employed to determine the validity of the seven instructional criteria.

Correlational analysis. First, a correlational matrix was constructed that included each of the 14 instructional level scores (seven criteria x two basal series) and the raw scores on the two

achievement tests. Correlations ranged from +.57 to +.95, reflecting the extent to which subjects' scores at the instructional level predict, or are valid, with respect to subjects' scores on the standardized achievement tests. Of 28 correlations (14 instructional level scores x 2 achievement test scores), 23 were greater than +.80. Averaged within instructional criteria, the mean correlations for Criterion 1 through Criterion 7 were +.93, +.88, +.62, +.85, +.85, +.86, and +.90, respectively. Correlations, then, for all of the criteria except for Criterion 3 were high and similar to each other.[1]

Congruency analyses. Two congruency analyses explored the extent of agreement between instructional level scores and three criterion measures. The criterion measures were (a) teachers' actual level of placements of subjects in the Ginn 720 series, (b) subjects' performance on the WI test, and (c) subjects' performance on the PC test. The first of these analyses examined whether subjects' reading levels, defined by each of the instructional criteria, were the same as, higher, or lower than subjects' reading levels denoted by each of the three criterion measures. Reading levels designated by instructional criteria were perceived as in agreement with teacher placements when instructional level scores fell within a range of two consecutive texts in the Ginn 720 series (-1 level $\leq x \leq$ + 1 level), or within an average of .88 grade levels. An instructional score was considered to be congruent with the two achievement tests when the instructional score was within 1.0 grade levels. Correlated *t* tests applied to the differences between instructional level scores and each of the three criterion measures constituted the second congruency analysis.

Table 3 displays the percentages of subjects placed high, low, and accurately with respect to teacher placements. Employing Criteria 4 through 7, the instructional scores placed similar percentages of subjects high, low, and accurately. Across the four performance standards, an average of 64.50% of the subjects were placed correctly, 17.00% were placed low, and 18.50% were placed high. Using Criterion 2, the extent of agreement was proportionately similar; however, a smaller percentage was placed correctly (53.00%) and greater percentages of subjects were placed high (29.00%) and low (18.00%). Instructional Criterion 3 placed low a relatively large percentage of subjects (58.00%) and Criterion 1 placed high a comparatively large percentage of subjects (50.00%).

Insert Table 3 about here

Correlated t tests corroborated this pattern of congruency for the different instructional criteria. For Criteria 1 and 2, the difference between the instructional scores and the teacher placements was statistically significant, $t(89) = 8.42$, $p = .000$ for Criterion 1 (mean difference = 1.87 levels) and $t(89) = 2.29$, $p = .000$ for Criterion 2 (mean difference = .54 levels). For Criterion 3 the difference also was statistically significant, $t(89) = 7.72$, $p = .000$. This time, however, the teacher placements were higher than the instructional scores (mean difference = 2.32 levels). For Criteria 4-7, there were no statistically significant differences.

The degree of congruency between the instructional level scores in both basal series and the PC and WI tests also were examined. Each

instructional level score was converted to its corresponding readability grade score (see Table 1). The readability grade score for each instructional criterion then was compared to both the WI and PC grade equivalency scores for every student to determine the percentages of students placed high, low, and accurately by each instructional criterion. Therefore, there were four combinations of congruency percentages and four series of correlated *t* tests: Ginn 720 series instructional grade scores with PC and WI grade scores, and Scott-Foresman Unlimited instructional grade scores with PC and WI grade scores.

The average percentages across these four combinations are presented in Table 4. The extent of congruency was similar for Criteria 4-7, with an average of 51.39% of students placed the same, 10.18% placed high, and 38.43% placed low. Criterion 2 placed correct a similar percentage (51.50%) with a more even distribution between low (21.50%) and high (26.50%) placements. Criterion 3 placed low a large percentage of students (60.25% placed low, 38.00% placed the same, and 1.00% placed high), while Criterion 1 placed high a large percentage of students (43.25% placed high, 11.25% placed low, 44.75% placed the same).

Insert Table 4 about here

Again, correlated *t* tests corroborated this pattern of congruency for different instructional criteria. For Criteria 1 and 3, the difference between the instructional grade scores and achievement test grade scores always was statistically significant for Criterion 1, $t(91) \leq 3.55$, $p = .001$ and for Criterion 3, $t(91) \leq 5.33$, $p = .000$. Criterion 1 placed students high by an average .55 levels and Criterion

3 placed students low by an average 1.29 levels, with respect to standardized test performance. The average difference was the smallest for Criterion 2 (.11 levels).

Discussion

The purpose of this investigation was to explore the reliability and validity of the following prominent IRI procedures: (a) choosing a 95% word recognition accuracy standard for determining instructional level; (b) arbitrarily selecting a passage to represent the difficulty level of a basal reader; and (c) employing one-level floors and ceilings. Findings of this study support the technical adequacy of one of these procedures, but question the adequacy of the remaining two.

Results support the use of the traditional, IRI standard of 95% for accuracy of word recognition. This standard of instructional level, as well as several other criteria used in informal reading assessment, exhibit validity with respect to standardized achievement tests. As evidence of this validity, correlations between instructional level scores and achievement test raw scores were high and statistically significant, except when Criterion 3 was employed. Criterion 3 was the level at which a student read at 100 wpm with 0-2 errors. This criterion, the most stringent, placed many students at low reading levels, failing to discriminate effectively among readers with different skills and resulting in lower correlations with achievement tests.

Two congruency analyses supplemented the correlational examination of the validity of IRI instructional performance standards. These analyses were: (a) the percentages of students placed, low, high, and the same with respect to criterion measures, and (b) correlated

t tests on the difference between the instructional level scores and the scores generated by criterion measures. These congruency analyses revealed that, despite its high correlations with the standardized tests, Criterion 1 yielded instructional level scores that did not agree well with either of the criterion measures, teacher placements, or the standardized tests. Criterion 3, which resulted in the lowest correlations with standardized tests, also produced instructional level scores that agreed poorly with both criterion measures.

To determine the acceptability of an instructional criterion, the following arbitrary standard was adopted. It had to produce scores that resulted in (a) correlations with standardized achievement tests of at least +.80; (b) at least 50.00% congruency with teacher placements and standardized tests; and (c) an average difference of no more than one-half level between instructional level scores and teacher placements and standardized tests. Given this standard of acceptability, Criteria 2, 4, 6, and 7 appear acceptable. Criterion 2 is 70 + wpm with 10 or fewer errors (86% accuracy). Criterion 4 is 95% accuracy, the traditional IRI instructional criterion. Criteria 6 and 7 employ different oral reading rates for primary (50 wpm) and intermediate (70 wpm) readers as they employ 95% and 95/85% accuracy, respectively. Any one of these four criteria demonstrates strong concurrent validity (as reflected in the correlations with standardized achievement tests) as well as good agreement with criterion measures. Each appears to be a good choice for use in an IRI.

Therefore, the external validity of several performance standards, including the popular IRI instructional performance standard, was

demonstrated in the present investigation. The strength of this conclusion, however, is tempered in light of two deviations from standard IRI procedure. First, in contrast to the typical one-level ceiling, a two-level ceiling was employed to determine instructional levels. A second deviation, also relevant to the remaining discussion, is that reading performance was timed in this study and students were stopped at the completion of 60 seconds.

With respect to the two other commonly employed IRI procedures, results of the present study question the typical passage selection procedure as well as the use of one-level ceilings and floors. First, for over one-half of the 19 books employed in the investigation, adequate readability representation was not achieved until 10 or more passages were sampled. Therefore, the common practice of arbitrarily selecting passages from a book to represent the difficulty of the material in that text appears inadequate, and may jeopardize the confidence with which educators can interpret IRI results.

Second, despite the use of representative passages that, in fact, did increase in difficulty within each reading series, students' performances did not necessarily weaken as a function of this increasing difficulty. An average of only one-half to three-quarters of mean performance scores decreased on adjacent passages. Additionally, for an average of over one-half of the subjects, (a) performance standards were met at levels higher than a level that the student already had failed, and/or (b) the standards were not met at levels lower than one at which the student had succeeded. These findings seriously question the assumption often held by advocates of IRIs that a student's

performance is consistently adequate below a one-level floor or that his/her performance is consistently inadequate above a one-level ceiling. To proceed on the basis of such an assumption may produce inaccurate estimates of pupils' instructional levels.

The findings of this study thus suggest that IRI procedures for selecting passages from basal texts and for sampling pupils' performance at instructional levels may have a negative effect on current educational practice. Alternate approaches to current procedures include: (a) identifying representative passages with readability formulae instead of employing arbitrarily selected passages to represent a text's difficulty level, and (b) requiring students to read representative passages from each level of a text rather than using a floor/ceiling approach. These alternate procedures may reduce error and may possess greater technical adequacy than current practice; however, they may reduce dramatically IRIs' appeal to practitioners. Curriculum-based IRIs seem to be popular as an informal assessment procedure because of the ease with which they can be created within any curriculum and then implemented. Relatively elaborate procedures for creating and administering curriculum-based IRIs may make them infeasible for classroom use.

We believe that another methodological option combines logistical feasibility with a capacity to sample both reading materials and pupils' competencies with greater validity. Epstein (1980) has suggested that sampling over occasions and over test forms is a widely ignored method for reducing measurement error and for increasing the likelihood of replicable findings. Based on this premise, an alternate strategy consists of creating parallel forms of IRIs, administering them on

consecutive days, and then aggregating pupils' reading performances over days or continuing administrations until results agree on at least two consecutive days. By testing over alternate forms, error stemming from nonrepresentative passages would be reduced because each day new passages would be employed; by assessing over occasions, error resulting from transitory student, examiner, situational, and procedural characteristics in testing also would be diminished. Additionally, by more stringently demanding agreement in results on at least two consecutive days or by aggregating performance over days to determine results, this procedure might reduce error that stems from the lack of consistency in the deterioration of student performance through a series of passages of increasing difficulty. For example, Lovitt and Hansen's (1976) data revealed that a student's performance did not consistently worsen as a function of increasingly more difficult passages on any one day. Yet, when averaged over five days, the student's performance did progress more consistently through the passages. While these procedures may be more time consuming than current practices, they still appear feasible and do not demand additional teacher training as other procedures might require.

References

Alper, T., Nowlin, L., Lemoine, K., Perine, M., & Bettencount, B. The rated assessment of academic skills. Academic Therapy, 1973, 9, 151-164.

Armbruster, B. B., Stevens, R. J., & Rosenshine, B. Analyzing content coverage and emphasis: A study of three curricula and two tests (Technical Report No. 26). Urbana-Champaign: Center for the Study of Reading, University of Illinois, 1977.

Arnold, B. B., & Arnold, R. D. Measures and judgments of reading level for disabled readers. The Minnesota Reading Quarterly, 1966, 11(1), 9-15.

Beery, A., Barrett, T. C., & Powell, W. R. Elementary reading instruction. Boston: Allyn & Bacon, 1969.

Beldin, H. L. Informal reading testing: Historical review and review of the research. In W. Durr (Ed.), Reading difficulties: Diagnosis, correction and remediation. Newark, Del.: International Reading Association, 1970.

Bormuth, J. R. Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 1969, 4, 358-365.

Botel, M. A comparative study of the validity of the Botel Reading Inventory and selected standardized tests. International Reading Association, Conference Proceedings, Part I, 1968, 13, 722-727.

Bradley, M. M., & Ames, W. S. Readability parameters of basal readers. Journal of Reading Behavior, 1977, 11(2), 175-183.

Bush, C. L., & Huebner, M. H. Strategies for reading in the elementary school. London: MacMillan, 1970.

Cooper, J. L. The effect of adjustment of basal reading materials on reading achievement. Unpublished doctoral dissertation, Boston University, 1952.

Dale, E., & Chall, J. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20.

Eaton, M., & Lovitt, T. C. Achievement tests vs. direct and daily measurement. In G. Semb (Ed.), Behavior analysis and education. Lawrence, Kan.: University of Kansas, 1972.

Epstein, S. The stability of behavior: II. Implications for psychological research. American Psychologist, 1980, 35(9), 790-806.

Fitzgerald, G. G. Reliability of the Fry sampling procedure. Reading Research Quarterly, 1980, 15(4), 489-503.

Fuchs, D., & Balow, B. Formulating an informal reading inventory. Unpublished manuscript, 1974. (Available from Special Education Programs, University of Minnesota, Minneapolis, Minnesota 55455).

Ginn and Company. Reading 720. Lexington, Mass.: Ginn (Xerox Corp.), 1976.

Haughton, E. Aims-growing and sharing. In J. Jordan & L. Robbins (Eds.), Let's try doing something else kind of thing. Arlington, Virg.: The Council for Exceptional Children, 1972.

Jenkins, J. R., & Pany, D. Standardized achievement tests: How useful for special education? Exceptional Children, 1978, 44(6), 448-453.

Johnson, M. S., & Kress, R. A. Informal reading inventories. Newark, Del.: International Reading Association, 1969.

Kelley, D. Using an informal reading inventory to place children in instructional materials. In W. Durr (Ed.), Reading difficulties: Diagnosis, correction, and remediation. Newark, Del.: International Reading Association, 1970.

Kender, J. P. How useful are informal reading tests? In A. Beery, T. C. Barrett, & W. R. Powell (Eds.), Elementary reading instruction. Boston: Allyn & Bacon, 1969.

Lovitt, T. C., & Hansen, C. L. Round one - Placing the child in the right reader. Journal of Learning Disabilities, 1976, 6, 347-353.

Lowell, R. E. Problems in identifying reading levels with informal reading inventories. In W. Durr (Ed.), Reading difficulties: Diagnosis, correction, and remediation. Newark, Del.: International Reading Association, 1970.

Nunnally, J. C. Tests and measurement: Assessment and prediction. New York: McGraw-Hill, 1959.

Oliver, J., & Arnold, R. D. Comparing a standardization test, an informal inventory and teacher judgment on third grade reading. Reading Improvement, 1978, 15(1), 56-59.

Pikulski, J. A critical review: Informal reading inventories. The Reading Teacher, 1974, 28, 141-151.

Powell, W. K. Validity of the IRI reading levels. Elementary English, 1971, 48, 637-642.

Salvia, J., & Ysseldyke, J. E. Assessment in special and remedial education, (2nd ed.). Boston: Houghton-Mifflin, 1981.

Scott-Foresman Systems, Revised. Unlimited Series. Glenview, Ill.: Scott, Foresman & Co., 1976.

Smith, N. B. Graded selections for informal reading diagnosis. New York: New York University Press, 1959.

Spache, G. A new readability formula for primary grade materials. Elementary English, 1953, 53, 410-413.

Starlin, C. Evaluating and teaching reading to "irregular" kids. Iowa Perspective. Iowa Department of Public Instruction. Dec., 1979, 1-11.

Starlin, C., & Starlin, A. Guidelines for continuous decision making. Bemidji, Minn.: Unique Curriculums Unlimited, 1974.

Walker, H., & Lev, J. Statistical inference. New York: Holt, Rinehart & Winston, 1969.

White, O. R. A pragmatic approach to the description of progress in the single case. Unpublished doctoral dissertation, University of Oregon, 1971.

Woodcock, R. Woodcock reading mastery tests manual. Circle Pines, Minn.: American Guidance Service, 1973.

Ysseldyke, J. E. Psychoeducational assessment and decision making. In J. E. Ysseldyke & P. Mirkin (Eds.), Proceedings of the Minnesota roundtable conference on assessment of learning disabled children (Monograph No. 8). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Footnote

Douglas Fuchs is a Postdoctoral Associate at the Institute for Research on Learning Disabilities. He is now at Clark University, Worcester, Massachusetts.

[1]Differences between these correlations are judged without the benefit of statistical probability because the test available for determining differences between correlations calculated on the same sample limits inference only to groups identical to the observed sample (Walker & Lev, 1969).

Table 1

Level Numbers, Grade Levels, and Readability Information on Passages from Two Reading Series

Series Level Number	Grade Levels	$\bar{X}$ Readability Score Across Passage	N[a]	SD[b]	$\bar{X}$ Readability Scores of Two Selected Passages
Ginn 720					
3-4	PP-P	2.02	8	.098	2.01
5	1-1	2.21	5	.117	2.20
6	2-1	2.43	6	.196	2.43
7	2-2	3.17	13	.536	3.10
8	3-1	3.60	10	.468	3.66
9	3-2	4.11	6	.142	4.05
10	4	5.00	11	.476	5.00
11	5	5.38	10	.534	5.36
12	6	5.81	14	.392	5.75
13	7	6.00	13	.593	6.03
Scott-Foresman					
2-3	PP-P	2.57	9	.439	2.57
4	1	2.73	5	.156	2.77
5-6	2-1	2.87	10	.282	2.95
7-8	2-2	3.29	7	.293	3.30
9-10	3-1	3.64	9	.754	3.59
11-12	3-2	4.02	13	.520	3.94
13-15	4	4.89	5	.252	4.82
16-18	5	5.64	11	.525	5.70
19-21	6	6.04	13	.144	6.03

[a]Number of passages required to achieve representativeness.

[b]Standard deviation across passages.

Table 2

Differences in Readability Scores Between Each Consecutive Pair of Passages in the Ginn 720 and Scott-Foresman Series

	Publisher's Level Number	Difference in Mean	t Value	p-Value
Ginn 720	3-4 vs. 5	.19	-2.30	.050
	5 vs. 6	.22	-2.31	.050
	6 vs. 7	.74	-3.49	.003
	7 vs. 8	.43	-2.79	.011
	8 vs. 9	.51	-3.17	.009
	9 vs. 10	.89	-5.78	.000
	10 vs. 11	.38	-1.70	.107
	11 vs. 12	.43	-2.17	.045
	12 vs. 13	.19	- .78	.441
Scott-Foresman	2-3 vs. 4	.16	-1.32	.198
	4 vs. 5-6	.14	-1.25	.235
	5-6 vs. 7-8	.40	-3.04	.009
	7-8 vs. 9-10	.35	-1.22	.248
	9-10 vs. 11-12	.38	-1.29	.219
	11-12 vs. 13-15	.87	-4.92	.000
	13-15 vs. 16-18	.75	-3.98	.001
	16-18 vs. 19-21	.40	-1.93	.068

Table 3

Percentages of Students Placed Below, Above, and the Same as Teacher Placements by Each Instructional Criterion (N=89)[a]

Instructional Criterion	Placement by Curriculum-based Measures Compared to Teacher Placement		
	Below	Same	Above
7	15	69	16
6	19	65	14
5	23	63	15
4	21	61	18
3	58	39	3
2	18	53	29
1	3	47	50

[a]No placement was reported for two students.

Table 4

Percentages of Students Placed Below, Above, and the Same as Achievement Test Scores by Each Instructional Criterion (N=91)[a]

Instructional Criterion	Curriculum-based Grade Scores Compared to Achievement Test Grade Scores		
	Below	Same	Above
7	32.50	58.00	8.75
6	40.00	51.75	7.50
5	42.50	49.00	7.75
4	39.25	46.50	13.50
3	61.00	38.00	1.00
2	26.50	51.50	21.50
1	11.25	44.75	43.25

[a]Percentages are across reading series and across achievement tests (WI and PC).

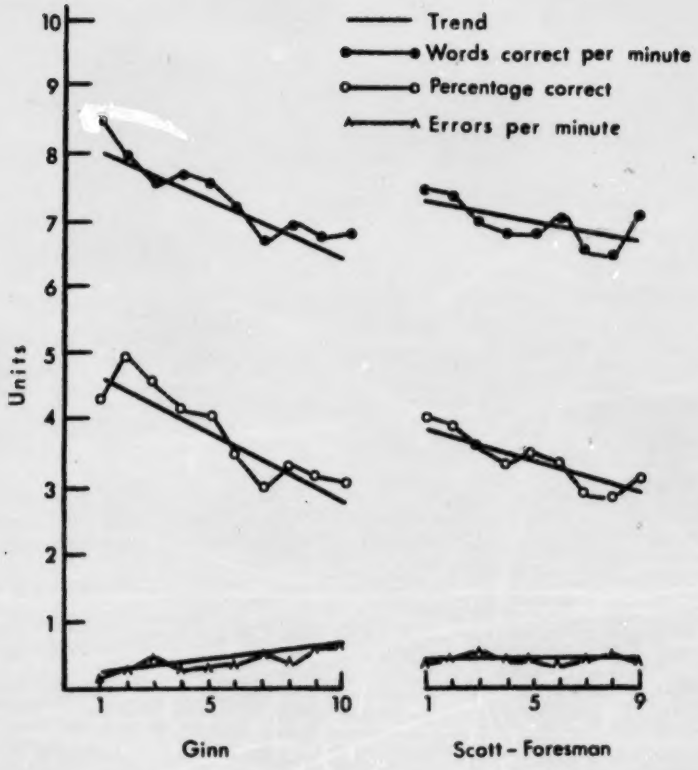

Figure 1. Number of words correct and errors per minute, and percentage correct in levels 1-10 of Ginn 720, and levels 1-9 of Scott-Foresman. Multiply units by 20.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota round-table conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). August, 1979.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled (Research Report No. 11). August, 1979.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDC's (Research Report No. 12). August, 1979.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers (Research Report No. 13). September, 1979.

Ysseldyke, J., & Algozzine, R. Perspectives on assessment of learning disabled students (Monograph No. 11). October, 1979.

Poland, S. F., Ysseldyke, J. E., Thurlow, M. L., & Mirkin, P. K. Current assessment and decision-making practices in school settings as reported by directors of special education (Research Report No. 14). November, 1979.

McGue, M., Shinn, M., & Ysseldyke, J. Validity of the Woodcock-Johnson psycho-educational battery with learning disabled students (Research Report No. 15). November, 1979.

Deno, S., Mirkin, P., & Shinn, M. Behavioral perspectives on the assessment of learning disabled children (Monograph No. 12). November, 1979.

Sutherland, J. H., Algozzine, B., Ysseldyke, J. E., & Young, S. What can I say after I say LD? (Research Report No. 16). December, 1979.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). December, 1979.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. The influence of test scores and naturally-occurring pupil characteristics on psycho-educational decision making with children (Research Report No. 17). December, 1979.

Algozzine, B., & Ysseldyke, J. E. Decision makers' prediction of students' academic difficulties as a function of referral information (Research Report No. 18). December, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information (Research Report No. 19). January, 1980.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). January, 1980.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 21). January, 1980.

Deno, S. L., Mirkin, P. K., & Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). January, 1980.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). January, 1980.

Deno, S. L., Mirkin, P. K., Robinson, S., & Evans, P. Relationships among classroom observations of social adjustment and sociometric rating scales (Research Report No. 24). January, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Factors influential on the psychoeducational decisions reached by teams of educators (Research Report No. 25). February, 1980.

Ysseldyke, J. E., & Algozzine, B. Diagnostic decision making in individuals susceptible to biasing information presented in the referral case folder (Research Report No. 26). March, 1980.

Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.

Ysseldyke, J. E., Regan, R. R., & Schwartz, S. Z. The use of technically adequate tests in psychoeducational decision making (Research Report No. 28). April, 1980.

Richey, L., Potter, M., & Ysseldyke, J. Teachers' expectations for the siblings of learning disabled and non-learning disabled students: A pilot study (Research Report No. 29). May, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Instructional planning: Information collected by school psychologists vs. information considered useful by teachers (Research Report No. 30). June, 1980.

Algozzine, B., Webber, J., Campbell, M., Moore, S., & Gilliam, J. Classroom decision making as a function of diagnostic labels and perceived competence (Research Report No. 31). June, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer-simulated investigation (Research Report No. 32). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., & Richey, L. Psychoeducational assessment and decision making: Individual case studies (Research Report No. 33). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R., Potter, M., & Richey, L. Technical supplement for computer-simulated investigations of the psychoeducational assessment and decision-making process (Research Report No. 34). July, 1980.

Algozzine, B., Stevens, L., Costello, C., Beattie, J., & Schmid, R. Classroom perspectives of LD and other special education teachers (Research Report No. 35). July, 1980.

Algozzine, B., Siders, J., Siders, J., & Beattie, J. Using assessment information to plan reading instructional programs: Error analysis and word attack skills (Monograph No. 14). July, 1980.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability (Research Report No. 36). July, 1980.

Algozzine, B., & Ysseldyke, J. E. An analysis of difference score reliabilities on three measures with a sample of low achieving youngsters (Research Report No. 37). August, 1980.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 38). August, 1980.

Richey, L. S., Ysseldyke, J., Potter, M., Regan, R. R., & Greener, J. Teachers' attitudes and expectations for siblings of learning disabled children (Research Report No. 39). August, 1980.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. L. (Eds.). A naturalistic investigation of special education team meetings (Research Report No. 40). August, 1980.

Meyers, B., Meyers, J., & Deno, S. Formative evaluation and teacher decision making: A follow-up investigation (Research Report No. 41). September, 1980.

Fuchs, D., Garwick, D. R., Featherstone, N., & Fuchs, L. S. On the determinants and prediction of handicapped children's differential test performance with familiar and unfamiliar examiners (Research Report No. 42). September, 1980.

Algozzine, B., & Stoller, L. Effects of labels and competence on teachers' attributions for a student (Research Report No. 43). September, 1980.

Ysseldyke, J. E., & Thurlow, M. L. (Eds.). The special education assessment and decision-making process: Seven case studies (Research Report No. 44). September, 1980.

Ysseldyke, J. E., Algozzine, B., Potter, M., & Regan, A. A descriptive study of students enrolled in a program for the severely learning disabled (Research Report No. 45). September, 1980.

Marston, D. Analysis of subtest scatter on the tests of cognitive ability from the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 46). October, 1980.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. Identifying children with learning disabilities: When is a discrepancy severe? (Research Report No. 47). November, 1980.

Fuchs, L., Tindal, J., & Deno, S. Effects of varying item domain and sample duration on technical characteristics of daily measures in reading (Research Report No. 48). January, 1981.

Marston, D., Lowry, L., Deno, S., & Mirkin, P. An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). January, 1981.

Marston, D., & Deno, S. The reliability of simple, direct measures of written expression (Research Report No. 50). January, 1981.

Epps, S., McGue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.

Evans, P. R., & Peham, M. A. S. Testing and measurement in occupational therapy: A review of current practice with special emphasis on the Southern California Sensory Integration Tests (Monograph No. 15). April, 1981.

Fuchs, L., Wesson, C., Tindal, G., & Mirkin, P. Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). June, 1981.

Fuchs, D., Featherstone, N., Garwick, D. R., & Fuchs, L. S. The importance of situational factors and task demands to handicapped children's test performance (Research Report No. 54). June, 1981.

Tindal, G., & Deno, S. L. Daily measurement of reading: Effects of varying the size of the item pool (Research Report No. 55). July, 1981.

Fuchs, L. S., & Deno, S. L. A comparison of teacher judgment, standardized tests, and curriculum-based approaches to reading placement (Research Report No. 56). August, 1981.

Fuchs, L., & Deno, S. The relationship between curriculum-based mastery measures and standardized achievement tests in reading (Research Report No. 57). August, 1981.

Christenson, S., Graden, J., Potter, M., & Ysseldyke, J. Current research on psychoeducational assessment and decision making: Implications for training and practice (Monograph No. 16). September, 1981.

Christenson, S., Ysseldyke, J., & Algozzine, B. Institutional constraints and external pressures influencing referral decisions (Research Report No. 58). October, 1981.

Fuchs, L., Fuchs, D., & Deno, S. Reliability and validity of curriculum-based informal reading inventories (Research Report No. 59). October, 1981.

Algozzine, B., Christenson, S., & Ysseldyke, J. Probabilities associated with the referral-to-placement process (Research Report No. 60). November, 1981.